Claims

What is claimed is:

1. A method for use in accordance with an audio-visual speech recognition system for improving a recognition performance thereof, comprising the steps of:

selecting between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment; and

decoding at least a portion of an input spoken utterance using the selected data model.

2. The method of claim 1, further comprising the step of storing the acoustic-only data model and the acoustic-visual data model in memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

3. The method of claim 1, wherein the model selection step is based on a likelihood ratio test.

4. The method of claim 3, wherein the model selection step further comprises selecting the acoustic-only data model when a result of the likelihood test is not greater than a threshold value.

5. The method of claim 3, wherein the model selection step further comprises selecting the acoustic-visual data model when a result of the likelihood test is not less than a threshold value.

6. The method of claim 5, wherein the threshold value is based on a cost associated with a recognition error.

7. The method of claim 3, wherein the likelihood ratio test is based on one or more observations of a given visual feature.

8. The method of claim 7, wherein the given visual feature is associated with the mouth region of a speaker of the input utterance.

9. The method of claim 1, wherein model selection is performed at a rate substantially equivalent to an observation rate associated with the audio-visual speech recognition system.

10. Apparatus for use in accordance with an audio-visual speech recognition system for improving a recognition performance thereof, the apparatus comprising:

a memory; and

at least one processor coupled to the memory and operative to: (i) select between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment; and (ii) decode at least a portion of an input spoken utterance using the selected data model.

11. The apparatus of claim 10, wherein the acoustic-only data model and the acoustic-visual data model are stored in the memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

12. The apparatus of claim 10, wherein the model selection operation is based on a likelihood ratio test.

13. The apparatus of claim 12, wherein the model selection operation further comprises selecting the acoustic-only data model when a result of the likelihood test is not greater than a threshold value.

14. The apparatus of claim 12, wherein the model selection operation further comprises selecting the acoustic-visual data model when a result of the likelihood test is not less than a threshold value.

15. The apparatus of claim 14, wherein the threshold value is based on a cost associated with a recognition error.

16. The apparatus of claim 12, wherein the likelihood ratio test is based on one or more observations of a given visual feature.

17. The apparatus of claim 16, wherein the given visual feature is associated with the mouth region of a speaker of the input utterance.

18. The apparatus of claim 10, wherein model selection is performed at a rate substantially equivalent to an observation rate associated with the audio-visual speech recognition system.

19. An article of manufacture for use in accordance with an audio-visual speech recognition system for improving a recognition performance thereof, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

selecting between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment; and

decoding at least a portion of an input spoken utterance using the selected data model.

20. The article of claim 19, further comprising the step of storing the acoustic-only data model and the acoustic-visual data model in memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

21. An audio-visual speech recognition system, comprising:

a memory; and

at least one processor coupled to the memory and operative to: (i) select between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment; and (ii) decode at least a portion of an input spoken utterance using the selected data model, wherein the acoustic-only data model and the acoustic-visual data model are stored in the memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

22. A method for use in accordance with a speech recognition system for improving a recognition performance thereof, comprising the steps of:

selecting for a given frame between a first data model and at least a second data model based on a given condition; and

decoding at least a portion of an input spoken utterance for the given frame using the selected data model.